ERNE

the EDRN Resource
Network Exchange for
Specimen Information

Steve Hughes/NASA JPL
Don Johnsey/NCI
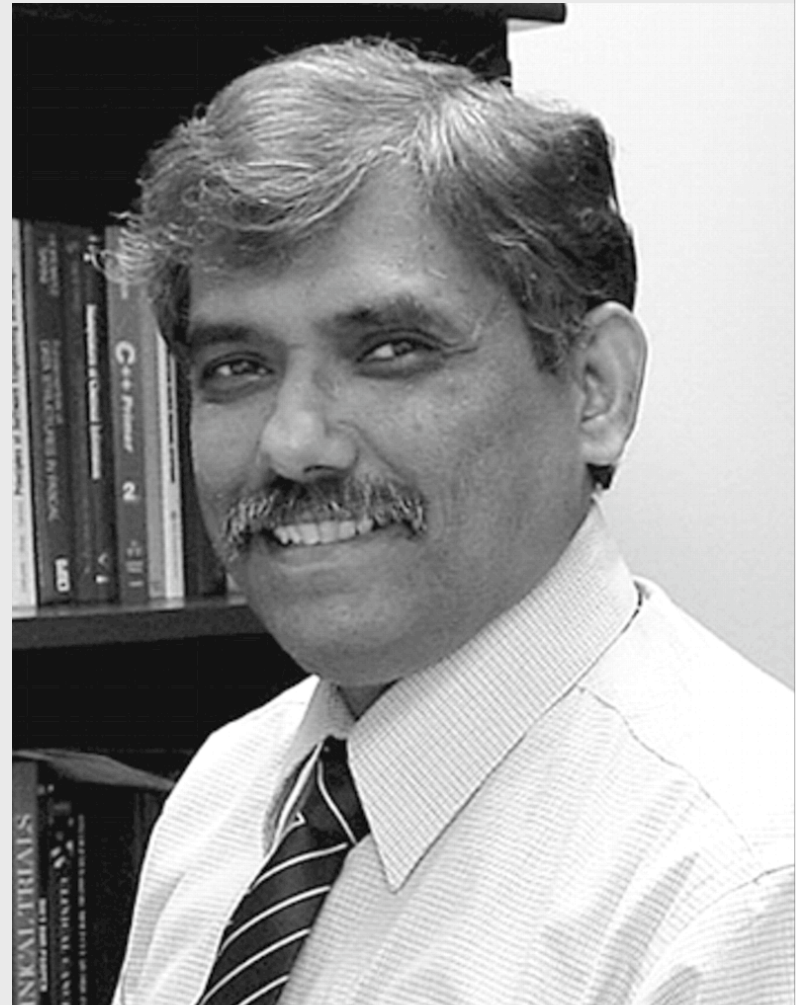**Dan Crichton/NASA JPL**
Sean Kelly/NASA JPL

Overview

- EDRN

- ERNE

- caTissue Integration

EDRN

- Early Detection Research Network

  - Sudhir Srivastava

  - Chief, Cancer Biomarkers Research Group

- Run by the Division of Cancer prevention

- National Cancer Institute

- National Institutes of Health

- Brings together dozens of institutions

- Accelerates biomarker research

  - Develops and tests promising markers

  - Evaluates promising technologies

  - Disseminates information rapidly

- Collaboration across disciplines:

  - Molecular biology and genetics

  - Clinical oncology

  - Computer science

  - Public health

  - Clinical application

Structure of EDRN
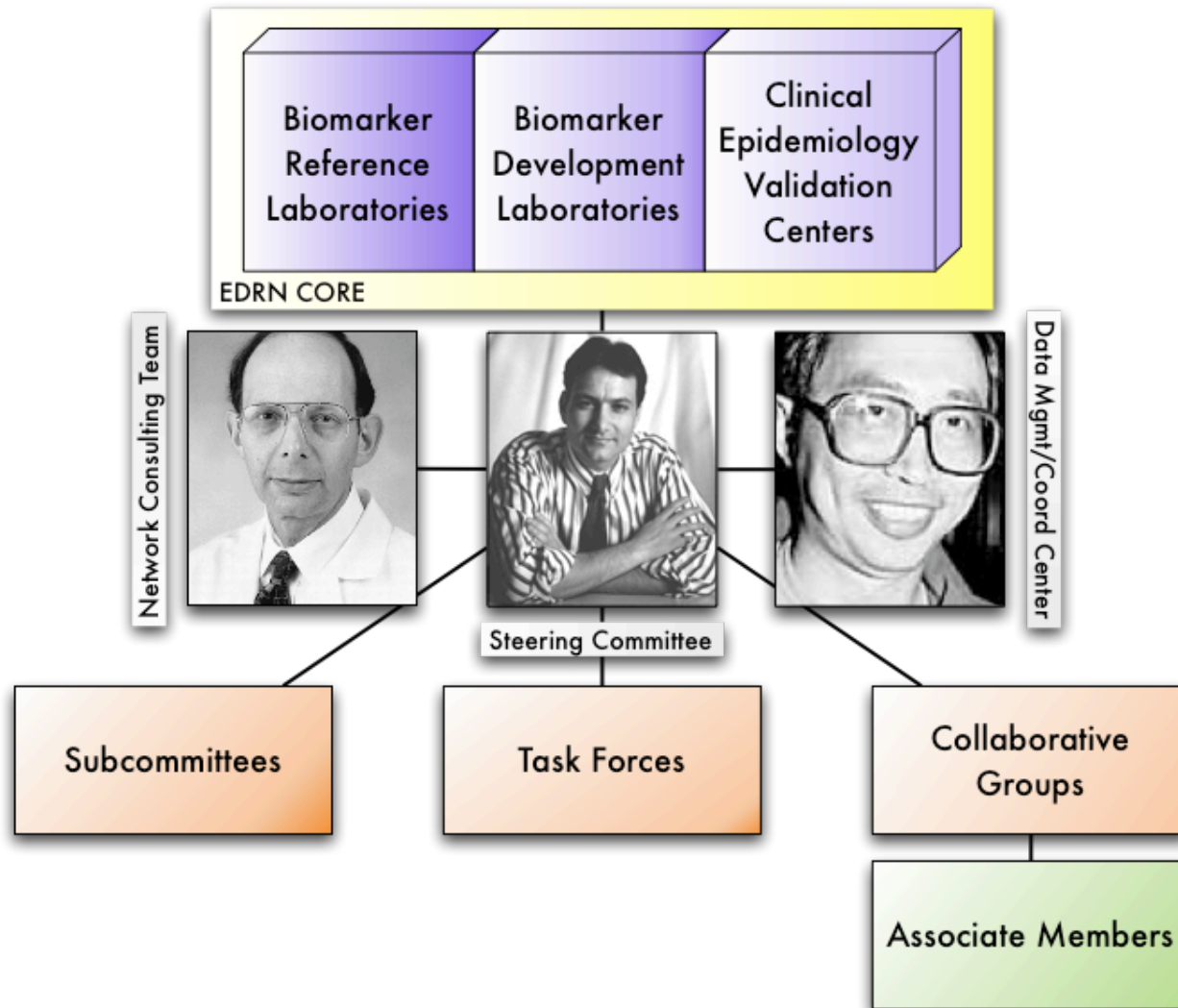
- Biomarker Development Laboratories

  - Develop new biomarkers

  - Characterize and refine existing markers

- Biomarker Reference Laboratories

  - Serves as EDRN resource

    - For clinical and lab validation of markers

  - Develops technology, QC, etc. of markers

- Clinical Epidemiology / Validation Centers

  - Conducts research into clinical application of markers

- Data Management/Coordinating Center

  - Provides logistics, statistical, computation support

  - Develops common EDRN database

- Informatics Center at NASA/JPL

  - Responsible for researching and developing informatics infrastructure and technologies for EDRN

- Steering Committee

  - Provides overall direction

- Collaborative Groups

  - Focus on specific organs

ERNE

- EDRN Resource Network Exchange
  - Creates virtual specimen database
  - Leverages NASA-developed software

- The problem

  - Each EDRN site collects specimens

  - Each EDRN site would like to share their collection

  - Each EDRN site catalogs their collection

But in different ways

- EDRN needed

  - A *virtual* specimen catalog

  - That was *non-intrusive*

    - Did not modify existing catalogs

    - Did not alter existing procedures

NASA's OODT

- Object-Oriented Data Technology

  - Unifies disparate resources

  - Metadata-based framework for data management

  - A set of distributed software components that communicate using a common language
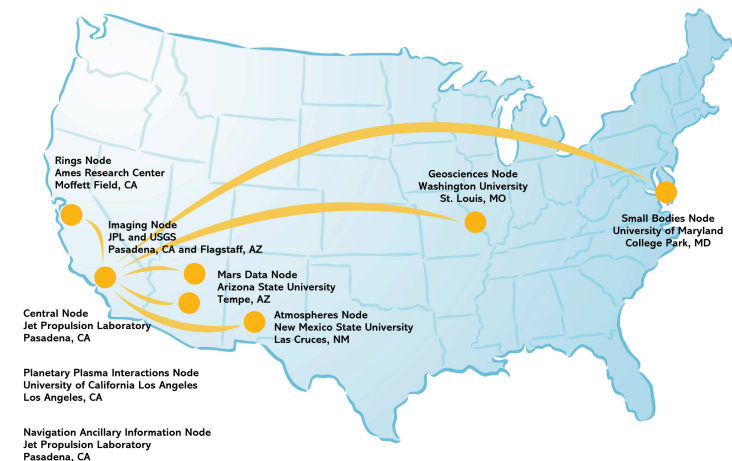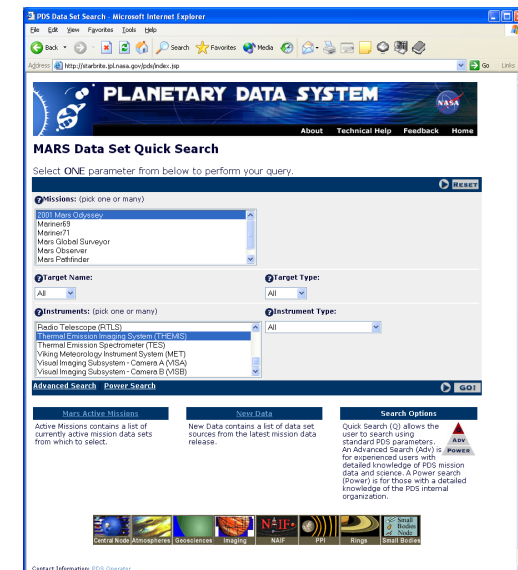
- OODT (*Object Oriented Data Technology*)

  - Component-based architecture developed in Java

  - Originally developed for

    - Astrophysical

    - Planetary

    - Atmospheric data management

- Core services (each distributable)

  - **Product service** retrieves data products from system-specific sources

  - **Profile service** locates data and resources using metadata search

  - **Catalog/archive service** ingests and processes incoming data products

- Deployed to

  - NASA's Planetary Data System which includes all solar system projects

  - NASA QuickSCAT satellite earth science mission

  - NASA SeaWinds earth science mission

  - NASA James Web Telescope for astrophysics

  - Orbiting Carbon Observatory (coming soon)

Planetary Data System

- Official NASA "Active" Archive for solar system mission data

- 9 Discipline Nodes with data geographically distributed at discipline sites (mostly universities)

- Uses a PDS Domain Information Model

- OODT deployed nationally at all sites

Rings Node
Ames Research Center
Moffett Field, CA

Imaging Node
JPL and USGS
Pasadena, CA and Flagstaff, AZ

Central Node
Jet Propulsion Laboratory
Pasadena, CA

Planetary Plasma Interactions Node
University of California Los Angeles
Los Angeles, CA

Navigation Ancillary Information Node
Jet Propulsion Laboratory
Pasadena, CA

Mars Data Node
Arizona State University
Tempe, AZ

Atmospheres Node
New Mexico State University
Las Cruces, NM

Geosciences Node
Washington University
St. Louis, MO

Small Bodies Node
University of Maryland
College Park, MD

Discovering/Correlating

- Photometer data

- Magnetometer data

- Accelerometer data

- Spectroradiometer

- Photography

- Reflectometer

- Interferometer

- Temperature

- Radiography

- Anemometer

- *And more,* from multiple planets, stars, and other targets!

- And deployed to EDRN

  - The *same core* components

  - The *same* OODT!

  - Just a different *domain information model* (e.g., different common data elements)

  - Deployed to 9 sites (going to 15 in 2006)

- ERNE uses the *product service*

- But ERNE needs common data elements

- CDEs

  - Define a core set of EDRN specimen CDEs

  - Define the vocabulary for making a specimen query

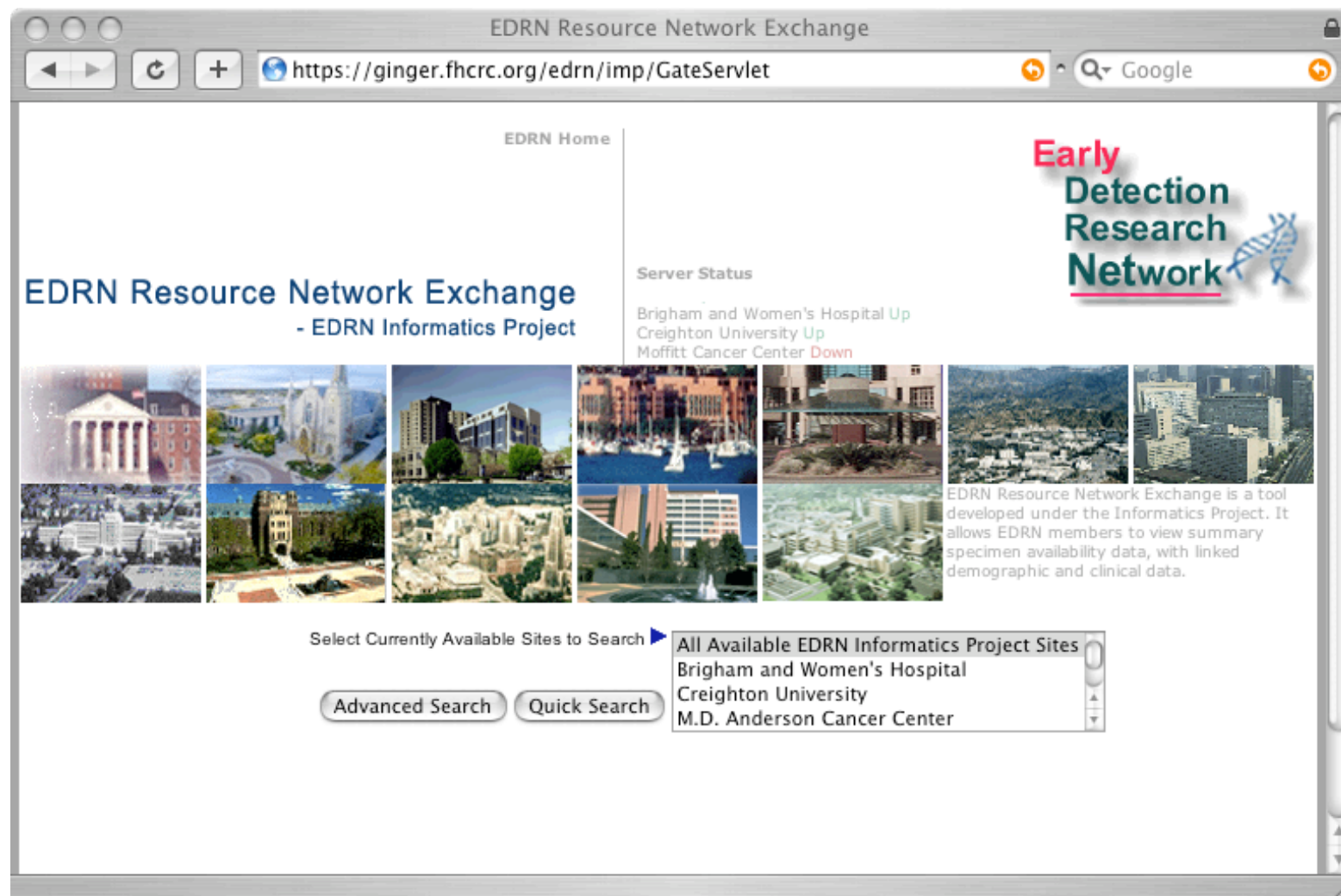 - Define the vocabulary for a response to such a query

ERNE Architecture
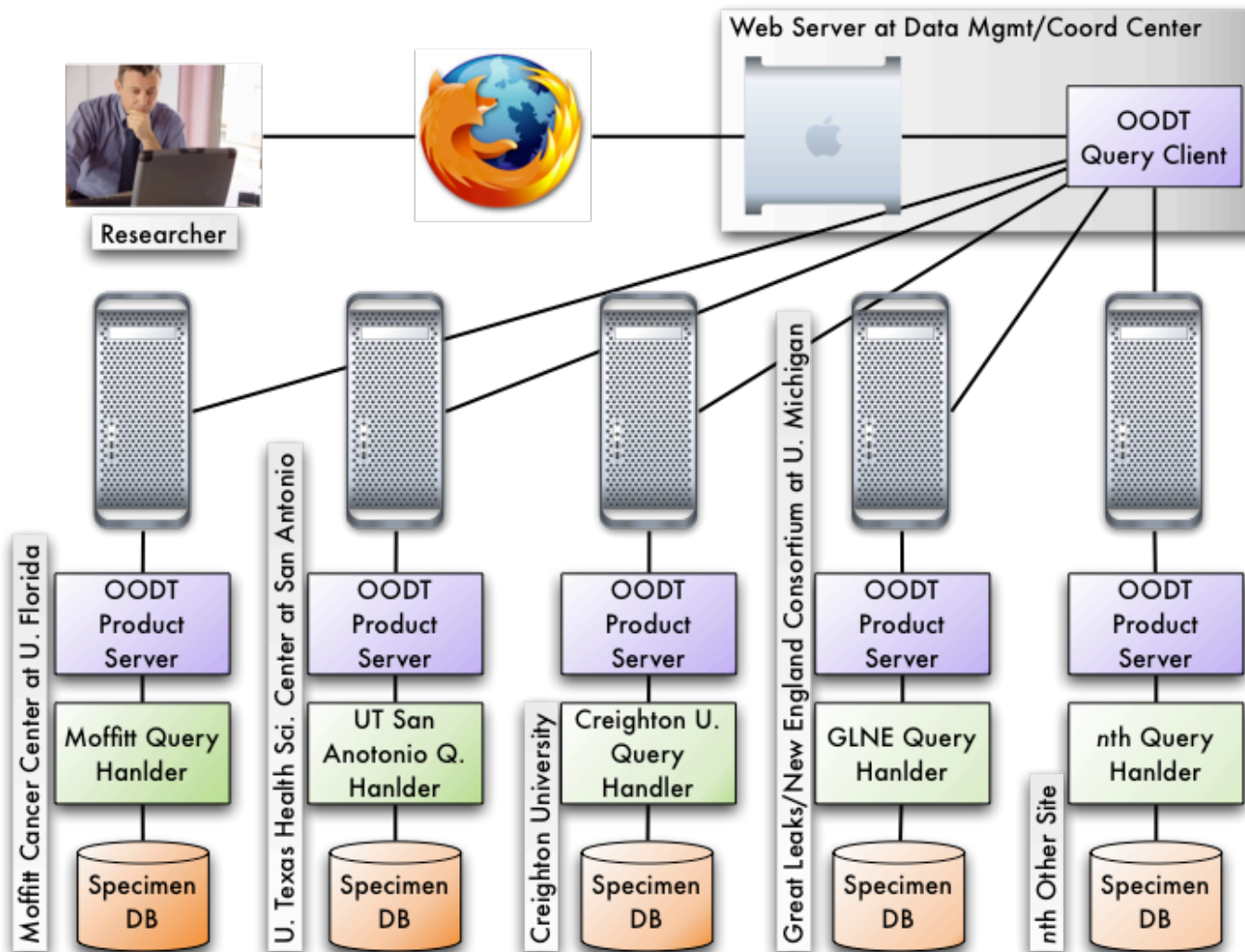
- Researcher uses a preferred web browser

  - Visits EDRN Secure Web Site

  - Clicks link to ERNE

- ERNE Welcome Page

  - Shows sites that are up

  - May search all or a subset of sites

ERNE Welcome Page

- Quick search

    - Pick specimen source

        - Blood, tissue, urine, etc.

    - Storage

        - DNA, frozen tissue, buffy coat, etc.

    - Participant status (w/ or w/o cancer)

Quick Search

- Advanced search

  - = quick search + final storage

  - + participant demographics

  - + individual cancer characteristics

  - + histology

  - + relatives cancer status, and more

EDRN Resource Network Exchange

← → C + https://ginger.fhcrc.org/edrn/imp/query.jsp?user=skelly&s=3&p=1&pas⟳ ^ Q▼ Google

EDRN Home

EDRN Resource Network Exchange
    - EDRN Informatics Project

Sites Selected:

Brigham and Women's Hospital
Creighton University
M. D. Anderson

**Early
Detection
Research
Network**

Listed specimens are available at sites selected. The fields with ✱ sign are required.

Specimen Source ✱ [Blood ⬍]   Participant Cancer Status ✱ [Participant With Cancer ⬍]

Characteristics of Specimen:

Specimen Stored ✱ [Lymphocytes ⬍]   Final Storage [All ⬍]

Demographics:

Gender: Hispanic/Latino Origin:
(●) All   ( ) Male   ( ) Female   (●) All   ( ) Hispanic/Latino   ( ) Not Hispanic/Latino

Race:
[✓] All   [ ] White   [ ] Black or African American
[ ] American Indian or Alaskan Native   [ ] Asian   [ ] Native Hawaiian/Pacific Islande

History of Regular Smoking: (●) All   ( ) Yes   ( ) No

Characteristics of Individuals with Cancer:

Cancer Sites:
[✓] All   [ ] Bladder   [ ] Bone   [ ] Brain
[ ] Breast(f)   [ ] Breast(m)   [ ] Cervix   [ ] Colon
[ ] Endometrium   [ ] Esophagus   [ ] H & N, floor of mouth   [ ] H & N, gum
[ ] H & N, hypopharynx   [ ] H & N, lip   [ ] H & N, nasal, ear, sinuses   [ ] H & N, nasopharynx
[ ] H & N, oropharynx   [ ] H & N, other   [ ] H & N, other mouth   [ ] H & N, tongue
[ ] Kidney   [ ] Leukemia   [ ] Liver   [ ] Lung
[ ] Lymphoma   [ ] Ovary   [ ] Pancreas   [ ] Prostate
[ ] Rectum   [ ] Skin   [ ] Stomach   [ ] Testes
[ ] Thyroid   [ ] Uterine   [ ] Uterus (unspec.)   [ ] Vaginal

Histology Classification:
[✓] All   [ ] Invasive Tumor   [ ] Pre-invasive Neoplasia   [ ] Hyperplasia
[ ] Other   [ ] Normal   [ ] Indeterminate

Specimen Collection Period:
From [All Prediagnosis Period ⬍]   To [All Postdiagnosis Period ⬍]   ◄——From——————————To——►
                                                                                    Prediagnosis   Diagnosis   Postdiagnosis

Age at Cancer Diagnosis: From Age [0] Years Old   To Age [90] Years Old

Other:

Study Design:
[✓] All   [ ] Case/Control   [ ] Longitudinal   [ ] Convenience Sample
[ ] Nested Case/Control   [ ] Registry

Search Based on Cancer History of 1st Degree Relatives: ( ) Yes   ( ) No

Options for 1st Degree Relatives with Cancer:
( ) With No Cancer   ( ) With Any Cancer   ( ) With Any Specific Cancers

Cancer Sites of 1st Degree Relatives:
[✓] All   [ ] Bladder   [ ] Bone   [ ] Brain   [ ] Breast
[ ] Cervix   [ ] Colon   [ ] Esophagus   [ ] Head & Neck   [ ] Kidney
[ ] Leukemia   [ ] Liver   [ ] Lung   [ ] Lymphoma   [ ] Ovary
[ ] Pancreas   [ ] Prostate   [ ] Rectum   [ ] Skin   [ ] Stomach
[ ] Testes   [ ] Thyroid   [ ] Uterus   [ ] Vagina

(Query String) (Raw Data) (Reset) (Search)

- Upon clicking "Search"

  - Web server constructs OODT query object

OODT Query

- `http://oodt.jpl.nasa.gov/edm-query/`

- Triples of

  - Data element (SPECIMEN_TYPE, …)

  - Relational operator (=, ≤, ≥, ≅, …)

  - Literal value (4, -39.63, saturn, …)

- Linked with boolean operators

  - AND

  - OR

  - NOT

- For example:

  - SPECIMEN_TYPE=3 AND SPECIMEN_AMOUNT_REMAINING >=43.3

  - In the above:

    - Two triples, one boolean operator

- Query passed to OODT product client

  - http://oodt.jpl.nasa.gov/grid-product/

    - One per site

    - Run in tandem in separate threads

    - Concurrent query to active sites

    - Time out for a site response: 60 seconds

- OODT Product Server runs at each site

  - Product server delegates to one or more *query handlers*

  - Currently, one query handler per site

- Query handler is an *interface*

  - Concrete classes *implement* the interface

  - Their job:

    - Act as *translation layer* between CDEs and site's specimen database

- Each query handler developed using a *mapping*

  - Mapping defined by curators at sites

  - Tells how to go from CDEs to site databases and back

  - NASA/JPL developer uses mapping to create query handler

- To support mapping

  - DMCC developed mapping web app

  - Enables curators to tell

    - About database structure

    - Correspondence to CDEs

    - Other plain-text details

- Mapping serves as a *software specification*

  - NASA/JPL developer uses it to create unique software for each site

    - However, many common mappings are factored out for re-use

- Sites are welcome to develop their own query handlers (so far, one site has)

- Ultimately

  - CDE query becomes an SQL query

  - SQL rows become CDE rows

Presenting Results

- Summary page shows breakdown

  - By site

    - By demographics

- Includes contact information

  - To request specimens to be sent

Detail Page

EDRN Resource Network Exchange

https://ginger.fhcrc.org/edrn/imp/DetailServlet

EDRN Resource Network...   ERNE Interface to caTiss...

EDRN Home

EDRN Resource Network Exchange
- EDRN Informatics Project

Another Search

Early
Detection
Research
Network

Data in Details: Specimen # 1 - #100 (Creighton University, Protocol ID: 65)

| PptID	Smoking Hx	Cancer Site[1]	Cancer of Relative[2]	Age at Dx	Age at Collection	Collection Date	Quantity[3]	Available
95008116	No	n/a	Yes	6	1794	21779 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008207	No	n/a	No	6	1795	21784 mo postx	$2.0*10^6$ cells	Yes
Gender: Male Race: White								
95008058	No	n/a	Yes	6	1795	21784 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008058	No	n/a	Yes	6	1795	21784 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008058	No	n/a	Yes	6	1795	21784 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008165	Yes	n/a	No	4	1792	21778 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008736	No	Breast(f)	No	4	1800	21869 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008777	No	Breast(f)	Yes	3	1791	21777 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008777	No	Breast(f)	Yes	3	1791	21777 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008793	No	Breast(f)	Yes	3	1799	21867 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008454	No	Breast(f)	Yes	5	1794	21791 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008454	No	Breast(f)	Yes	5	1794	21791 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008454	No	Breast(f)	Yes	5	1794	21791 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008454	No	Breast(f)	Yes	5	1794	21791 mo postx	$1.0*10^6$ cells	Yes
Gender: Female Race: White								
95008454	No	Breast(f)	Yes	5	1794	21791 mo postx	$1.0*10^6$ cells	Yes

- Breaks down by participant

  - With details on availability

  - Quantity remaining

  - Other demographic details
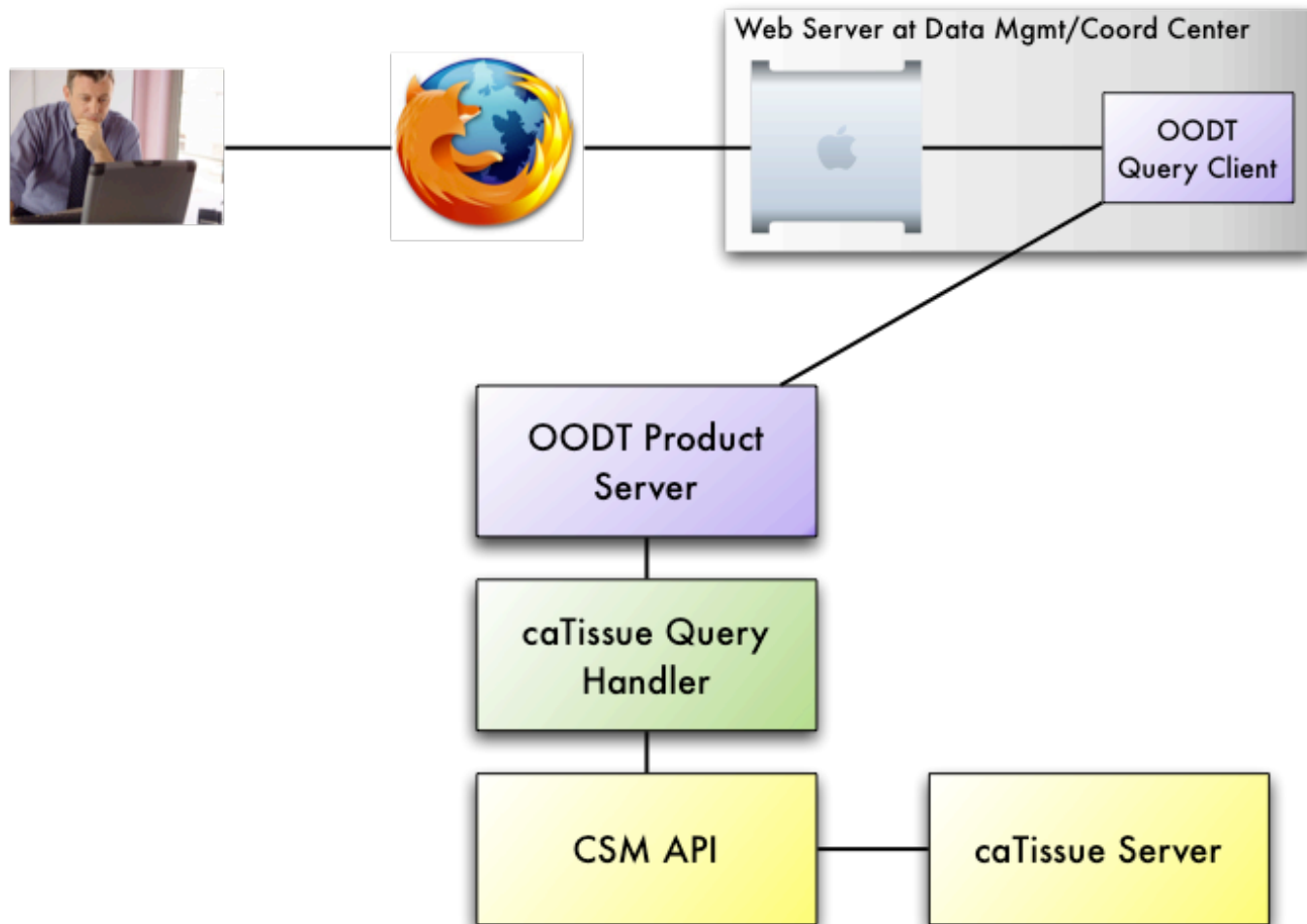
Architecture → Integration

- The query handler is the translation layer

  - Product client doesn't care what the query handler actually is

- Create even more query handlers!

FreezerWorks
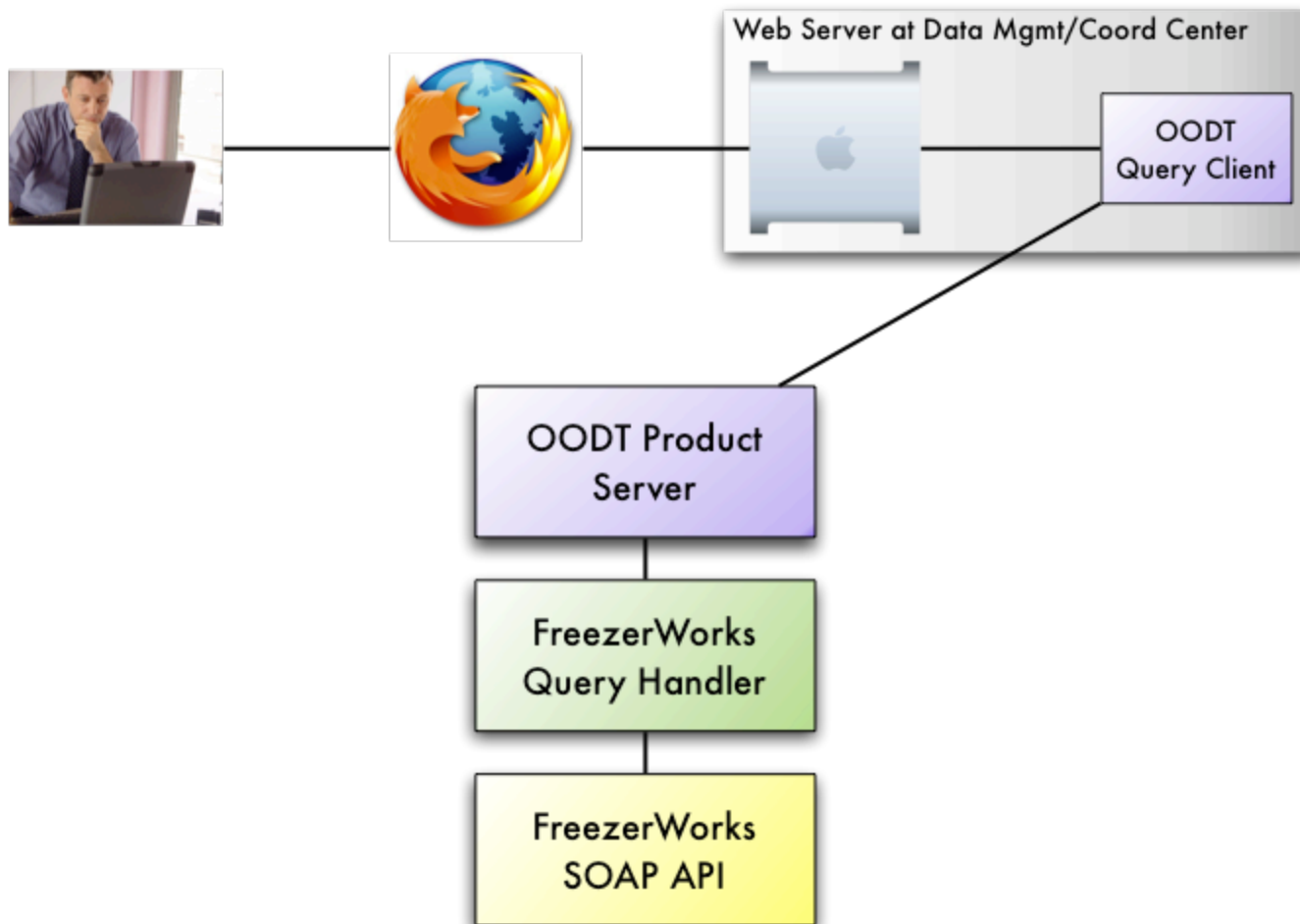
- Any FreezerWorks site becomes an ERNE site

    - Hypothetically

    - We haven't implemented this yet

caTissue

- Any caTissue site becomes an ERNE site

  - Again, hypothetical

  - But *currently in development*

Integrating caTissue→ERNE

- Get a test caTissue working

- Get a test CSM client working

- Develop ERNE→caTissue mapping

  - For querying for specimens

- Develop caTissue→ERNE mapping

  - For describing specimens

- NASA/JPL developer deployed caTissue 1.0 to local development environment

  - Tricky

  - But working in the end

- Developer created prototype query client using CSM API to caTissue

  - Demonstrated queries and retrieval of specimen data from external system

- Developer proposed caTissue/ERNE mapping

  - Available at Wiki:

    - http://oodt.jpl.nasa.gov/wiki/x/5AI

- Next steps

    - Finalize mapping

    - Develop caTissue/ERNE query handler

    - Develop installer

- Installer will ask

  - Location of caTissue system to use

  - User name and password to use

- End result

  - *caTissue system is now an ERNE site*

Expectations

- caTissue and ERNE officials will finalize mapping: 1 week

- OODT query handler development: 2 days

- Packaging and polish: 2 days

Summary

- EDRN

- ERNE

- caTissue Integration

- OODT

Questions?